- 25 -

THAT WHICH IS CLAIMED IS:

1. A computer-based method for screening a nucleic acid sequence for efficient translation in a predetermined host, said method comprising the steps of:

(a) providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(b) providing a substrate nucleic acid sequence;

(c) determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate;

(d) determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream from said first alignment of interest;

(e) determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest;

(f) successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

(h) detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle and correct phase through said substrate nucleic acid sequence indicating said substrate nucleic acid sequence is a candidate for efficient translation in said host.

- 26 -

2. A method according to claim 1, wherein said substrate nucleic acid sequence is heterologous to said host.

3. A method according to claim 2, wherein said substrate nucleic acid sequence is obtained from a different host than said ribosomal nucleic acid sequence.

4. A method according to claim 2, wherein said substrate nucleic acid sequence is synthetic.

5. A method according to claim 1, wherein said substrate nucleic acid sequence encodes a predetermined protein or peptide.

6. A method according to claim 1, wherein said detecting further comprises the step of determining the strength of said periodic signal.

7. A method according to claim 6, further comprising the step of:

(i) generating a quantitative indicator of translation efficiency from the strength of said periodic signal.

8. A method according to claim 6, further comprising the steps of:

(j) determining the sufficiency of said translation efficiency from said quantitative indicator; and then, in the absence of sufficient translation efficiency;

(k) replacing at least one base in said substrate nucleic acid sequence with a different base to produce a subsequent substrate nucleic acid sequence different from said previous substrate nucleic acid sequence, and encoding the same protein or peptide; and

(l) repeating steps (c) through (j) above with said subsequent substrate nucleic acid sequence.

9. A method according to claim 1, further comprising the steps of:

(j) determining the sufficiency of said translation efficiency from said quantitative indicator; and then, in the absence of sufficient translation efficiency;

(k) replacing at least one base in said substrate nucleic acid sequence with a different base to produce a subsequent substrate nucleic acid sequence different from said previous substrate nucleic acid sequence, and encoding the same protein or peptide; and

5   (l) repeating steps (c) through (j) above with said subsequent substrate nucleic acid sequence.

10. A method according to claim 8 or 9, wherein said step (l) further comprises repeating steps (c) through (k) until a nucleic acid sequence having
10   sufficient efficiency of translation in said host is identified.

11. A method according to claim 1, further comprising the steps of :

detecting a phase shift in said three base periodic cycle; and

determining the presence of a frame shift in said substrate nucleic acid
15   sequence from said phase shift, so that said substrate nucleic acid sequence remains a candidate for efficient translation in said host in the presence of said phase shift.

12. A method according to claim 1, wherein said host is a prokaryotic species.

20   13. A method according to claim 1, wherein said host is a eukaryotic species.

14. A method according to claim 1, further comprising the step of:

predicting the potential for proper translation of said substrate nucleic acid sequence in said host using the said already determined binding strengths of said
25   ribosomal nucleic acid sequence to said substrate nucleic acid sequence.

15. A method according to claim 1, wherein said step (g) of generating a binding strength pattern comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said
30   binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding

strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

16. A method according to claim 1, wherein said step (h) of detecting the presence or absence of a three-base binding strength periodic cycle comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

17. A method according to claim 1, wherein said step (f) of successively repeating steps (c) through (e) further comprising the step of:
calculating a cumulative energy for each said first, second, or third alignment of interest.

18. A method according to claim 1, further comprising the step of:
calculating a cumulative energy differential for a portion of said substrate nucleic acid sequence.

19. A method according to claim 1, further comprising the step of:
calculating a power spectrum magnitude of said binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.

20. A method according to claim 1, further comprising the step of:

calculating a mean binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.

5   21. A system for screening a nucleic acid sequence for efficient translation in a predetermined host, comprising:

(a) means for providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said

10   host:

(b) means for providing a substrate nucleic acid sequence;

(c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate:

15   (d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream from said first alignment of interest;

(e) means for determining a third binding strength of said ribosomal nucleic

20   acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest:

(f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step

25   (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence:

(g) means generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e): and

30   (h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength

- 30 -

cycle and correct phase through said substrate nucleic acid sequence indicating said substrate nucleic acid sequence is a candidate for efficient translation in said host.

22. A computer program product for screening a nucleic acid sequence for efficient translation in a predetermined host, said computer program product comprising a computer usable storage medium having computer readable program code means embodied in the medium, the computer readable program code means comprising:

(a) means for providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3` end of the 16S or 18S rRNA ribosomal subunit of said host;

(b) means for providing a substrate nucleic acid sequence;

(c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate;

(d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream from said first alignment of interest;

(e) means for determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest;

(f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

(h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle and correct phase through said substrate nucleic acid sequence indicating said

5   substrate nucleic acid sequence is a candidate for efficient translation in said host.

23. A computer-based method for screening a nucleic acid sequence for the presence of at least one coding sequence therein, said method comprising the steps of:

(a) providing a substrate nucleic acid sequence from a predetermined host;

10   (b) providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(c) determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

15   (d) determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

(e) determining a third binding strength of said ribosomal nucleic acid

20   sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third alignment of interest is two bases downstream from said first alignment of interest;

(f) successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is

25   three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

30   (h) detecting the presence or absence of a three-base binding strength periodic cycle and phase in at least one portion of the said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding

strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

24. A method according to claim 23, further comprising the step of:

5

(i) determining a start region and a stop region of said coding portion within said substrate nucleic acid sequence.

25. A method according to claim 23, wherein said detecting further comprises the step of determining the strength of said periodic signal.

10

26. A method according to claim 25, further comprising the step of:

(i) generating a quantitative indicator of translation efficiency from the strength of said periodic signal.

15

27. A method according to claim 23, further comprising the steps of :

detecting a phase shift in said three base periodic cycle; and

determining the presence of a frame shift in said substrate nucleic acid sequence from said phase shift, so that said substrate nucleic acid sequence remains a candidate for efficient translation in said host in the presence of said phase shift.

20

28. A method according to claim 23, wherein said host is a prokaryotic species.

29. A method according to claim 23, wherein said host is a eukaryotic species.

25

30. A method according to claim 23, further comprising the step of:

predicting the potential for proper translation of said substrate nucleic acid sequence in said host using the said already determined binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence.

30

31. A method according to claim 23, wherein said step (g) of generating a binding strength pattern comprises one or more calculating step(s), said calculating

step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding

5 strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

10   32. A method according to claim 23, wherein said step (h) of detecting the presence or absence of a three-base binding strength periodic cycle comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said

15 binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

20

       33. A method according to claim 23, wherein said step (f) of successively repeating steps (c) through (e) further comprising the step of:
       calculating a cumulative energy for each said first, second, or third alignment of interest.

25

       34. A method according to claim 23, further comprising the step of:
       calculating a cumulative energy differential for a portion of said substrate nucleic acid sequence.

30   35. A method according to claim 23, further comprising the step of:

- 34 -

calculating a power spectrum magnitude of said binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.

5   36. A system for screening a nucleic acid sequence for the presence of at least one coding sequence therein, comprising:

(a) means for providing a substrate nucleic acid sequence from a predetermined host;

(b) means for providing a ribosomal nucleic acid sequence, said ribosomal

10   nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3` end of the 16S or 18S rRNA ribosomal subunit of said host;

(c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

15   (d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

(e) means for determining a third binding strength of said ribosomal nucleic

20   acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third alignment of interest is two bases downstream from said first alignment of interest;

(f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step

25   (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e);

30   and

(h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase in at least one portion of the said substrate nucleic

acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

5

37. A computer program product for screening a nucleic acid sequence for the presence of at least one coding sequence therein, said computer program product comprising a computer usable storage medium having computer readable program code means embodied in the medium, the computer readable program code means

10 comprising:

(a) means for providing a substrate nucleic acid sequence from a predetermined host;

(b) means for providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the

15 first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

(d) means for determining a second binding strength of said ribosomal nucleic

20 acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

(e) means for determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest,

25 wherein said third alignment of interest is two bases downstream from said first alignment of interest;

(f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately

30 preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

(h) means for detecting the presence or absence of a three-base binding
5   strength periodic cycle and phase in at least one portion of the said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

10